

SADC Advanced Sampling Methods Workshop

International Standards for Household Survey Implementation

Project Overview

This repository contains comprehensive materials for a 5-day intensive training workshop on Advanced Sampling Methods for Household Surveys, designed for supervisors from 16 SADC member states' national statistics offices. The workshop integrates international best practices from Eurostat, World Bank, and OECD with SADC-specific contexts.

Workshop Details

- **Duration:** 5 days (Monday-Friday)
- **Daily Schedule:** 8:00 AM - 5:00 PM
- **Participants:** 50 supervisors from SADC national statistics offices
- **Instructor:** Harry (International Consultant with Eurostat/World Bank/OECD experience)
- **Location:** SADC Regional Training Center

Directory Structure

```
SADC_Sampling_Workshop/  
|  
|— 00-Setup/  
|   |— Installation_Guide.md  
|   |— R_Environment_Setup.R  
|   |— Package_Requirements.txt  
|   |— System_Check.R  
|  
|— 01-Data/  
|   |— household_survey_main_2024.csv  
|   |— household_roster_2024.csv  
|   |— enumeration_areas_master.csv  
|   |— auxiliary_census_2022.csv  
|   |— mobile_populations_sample.csv  
|   |— panel_rotation_cohort_2023.csv  
|   |— mixed_mode_responses.csv  
|   |— small_area_indicators.csv  
|   |— Metadata_household_survey_main.txt  
|   |— Metadata_household_roster.txt  
|   |— Metadata_enumeration_areas.txt
```

- | |— Metadata_auxiliary_census.txt
- | |— Metadata_mobile_populations.txt
- | |— Metadata_panel_rotation.txt
- | |— Metadata_mixed_mode.txt
- | |— Metadata_small_area.txt

|— 02-Scripts/

- | |— Script_1.1_Environment_Setup.R
- | |— Script_1.2_Data_Import_Quality.R
- | |— Script_1.3_Sampling_Frame_Construction.R
- | |— Script_1.4_Design_Effect_Calculation.R
- | |— Script_2.1_PPS_Selection.R
- | |— Script_2.2_Variance_Estimation_Taylor.R
- | |— Script_2.3_Small_Area_EBLUP.R
- | |— Script_2.4_Panel_Rotation_Weights.R
- | |— Script_3.1_Calibration_GREG.R
- | |— Script_3.2_Nonresponse_Adjustment.R
- | |— Script_3.3_Mobile_Population_TLS.R
- | |— Script_3.4_RDS_Implementation.R
- | |— Script_4.1_Mixed_Mode_Propensity.R
- | |— Script_4.2_Mode_Effect_Adjustment.R
- | |— Script_4.3_Quality_Indicators_ESS.R
- | |— Script_4.4_Total_Survey_Error.R
- | |— Script_5.1_Operational_Integration.R
- | |— Script_5.2_Documentation_Standards.R
- | |— Script_5.3_Dissemination_Tables.R
- | |— Script_5.4_Workshop_Synthesis.R
- | |— Validation_1.1.R
- | |— Validation_2.1.R
- | |— Validation_3.1.R
- | |— Validation_4.1.R
- | |— Validation_5.1.R

|— 03-Outputs/

- | |— design_diagnostics/
- | |— variance_estimates/
- | |— small_area_maps/
- | |— quality_reports/
- | |— final_weights/

|— 04-Presentations/

- | |— Day1_Foundations_International_Standards.Rmd
- | |— Day2_Implementing_International_Methodologies.Rmd
- | |— Day3_Advanced_Estimation_Techniques.Rmd

- | | — Day4_Special_Populations_Mixed_Mode.Rmd

- | | — Day5_Quality_Frameworks_Integration.Rmd

- | — 05-Exercises/

- | | — Exercise_1.1_Sampling_Frame_Assessment.md

- | | — Exercise_1.2_Design_Effect_Practical.md

- | | — Exercise_2.1_PPS_Implementation.md

- | | — Exercise_2.2_Variance_Comparison.md

- | | — Exercise_2.3_SAE_Model_Selection.md

- | | — Exercise_3.1_Calibration_Diagnostics.md

- | | — Exercise_3.2_Mobile_Population_Weights.md

- | | — Exercise_4.1_Mode_Effect_Analysis.md

- | | — Exercise_4.2_Quality_Framework_Application.md

- | | — Exercise_5.1_Complete_Survey_Design.md

- | — 06-Solutions/

- | | — Solution_1.1_Sampling_Frame.R

- | | — Solution_1.2_Design_Effect.R

- | | — Solution_2.1_PPS.R

- | | — Solution_2.2_Variance.R

- | | — Solution_2.3_SAE.R

- | | — Solution_3.1_Calibration.R

- | | — Solution_3.2_Mobile.R

- | | — Solution_4.1_Mode.R

- | | — Solution_4.2_Quality.R

- | | — Solution_5.1_Complete.R

- | — 07-Resources/

- | | — Bibliography.md

- | | — Quick_Reference_Cards.pdf

- | | — Formula_Compendium.md

- | | — Software_Comparison_Table.md

- | | — SADC_International_Standards_Comparison.md

- | | — Glossary_Terms.md

- | — 08-Harry-Journey/

- | | — Daily_Narratives.md

- | | — Participant_Interactions.md

- | | — Technical_Challenges_Log.md

- | | — Evening_Reflections.md

Technical Requirements

R Version and Core Packages

- R version 4.3.0 or higher
- RStudio 2023.06.0 or higher
- Rtools43 (Windows users)

Essential R Packages

```
r

# Survey statistics
- survey (4.2-1)
- sampling (2.10)
- sae (1.3)
- pps (1.0)

# Data manipulation
- tidyverse (2.0.0)
- data.table (1.14.8)
- janitor (2.2.0)

# Visualization
- ggplot2 (3.4.3)
- plotly (4.10.2)
- sf (1.0-14)
- tmap (3.3-4)

# Reporting
- xaringan (0.28)
- knitr (1.43)
- rmarkdown (2.24)
- kableExtra (1.3.4)

# Statistical methods
- lme4 (1.1-34)
- Matrix (1.6-1)
- MASS (7.3-60)
- boot (1.3-28)
```

Workshop Schedule

Day 1: Foundations and International Standards

- Module 1 (8-9 AM): Evolution of Sampling Theory and International Frameworks
- Module 2 (9-10 AM): SADC Statistical Systems and Harmonization Requirements
- Module 3 (10-11 AM): Sampling Frame Construction Following EU-SILC Standards
- Module 4 (11-12 PM): Design Effect Theory and Practical Calculations
- Module 5 (12-1 PM): Stratification Principles from LSMS Experience
- Module 6 (1-2 PM): Sample Size Determination with International Benchmarks
- Module 7 (2-3 PM): Documentation Standards (DDI, SDMX)
- Module 8 (3-4 PM): Introduction to R Survey Package Ecosystem

Day 2: Implementing International Methodologies

- Module 1 (8-9 AM): Two-Stage Sampling with PPS Implementation
- Module 2 (9-10 AM): Variance Estimation Using Taylor Linearization
- Module 3 (10-11 AM): Small Area Estimation Fundamentals (EBLUP)
- Module 4 (11-12 PM): Fay-Herriot Model Implementation
- Module 5 (12-1 PM): Rotating Panel Design (EU-SILC Approach)
- Module 6 (1-2 PM): Panel Weighting and Attrition Adjustment
- Module 7 (2-3 PM): Mobile Population Sampling (Time-Location)
- Module 8 (3-4 PM): Mixed-Mode Protocol Development

Day 3: Advanced Estimation Techniques

- Module 1 (8-9 AM): Calibration Using GREG Methodology
- Module 2 (9-10 AM): Raking and Distance Functions
- Module 3 (10-11 AM): Non-response Adjustment Models
- Module 4 (11-12 PM): Propensity Score Methods
- Module 5 (12-1 PM): Composite Estimation
- Module 6 (1-2 PM): Benchmarking and Coherence
- Module 7 (2-3 PM): Seasonal Adjustment in Surveys
- Module 8 (3-4 PM): Outlier Treatment Protocols

Day 4: Special Populations and Mixed-Mode Surveys

- Module 1 (8-9 AM): Hard-to-Reach Population Strategies
- Module 2 (9-10 AM): Respondent-Driven Sampling Implementation
- Module 3 (10-11 AM): Adaptive and Responsive Design
- Module 4 (11-12 PM): Web-First Sequential Mixed-Mode
- Module 5 (12-1 PM): Mode Effect Detection and Adjustment
- Module 6 (1-2 PM): Paradata Analysis for Quality
- Module 7 (2-3 PM): Cost-Quality Optimization Models
- Module 8 (3-4 PM): Real-Time Quality Monitoring Systems

Day 5: Quality Frameworks and Operational Integration

- Module 1 (8-9 AM): ESS Quality Framework Implementation
- Module 2 (9-10 AM): Total Survey Error Decomposition
- Module 3 (10-11 AM): Quality Indicators and Reporting
- Module 4 (11-12 PM): Process Quality Management
- Module 5 (12-1 PM): Integration with National Statistical Systems
- Module 6 (1-2 PM): Capacity Building Strategies
- Module 7 (2-3 PM): Dissemination and User Engagement
- Module 8 (3-4 PM): Workshop Synthesis and Action Plans

Quality Assurance Framework

Validation Checkpoints

1. **Design Phase:** Frame coverage, stratification efficiency, sample allocation
2. **Implementation Phase:** Response rates, interviewer effects, mode compliance
3. **Estimation Phase:** Weight diagnostics, variance stability, coherence checks
4. **Dissemination Phase:** Metadata completeness, accessibility standards, reproducibility

Performance Indicators

- Design Effect (DEFF): Target range 1.5-3.0
- Coefficient of Variation (CV): <5% for national estimates, <20% for domains
- Response Rate: Minimum 70% for face-to-face, 50% for mixed-mode
- Item Non-response: <10% for key variables
- Editing Rate: <5% automatic edits, <2% manual interventions

International Standards References

- Eurostat (2021): EU-SILC Methodology Manual, Version 2021.1
- World Bank (2020): LSMS Guidebook, Updated Edition
- OECD (2019): PIAAC Technical Standards and Guidelines
- UN Statistical Division (2022): Handbook on Household Surveys, Revised Edition
- ILO (2020): Labour Force Survey Manual
- IMF (2021): Data Quality Assessment Framework

SADC Context Adaptations

- Multi-lingual questionnaire protocols (English, Portuguese, French)
- Urban/rural stratification reflecting SADC development patterns
- Migration and cross-border population considerations
- Informal settlement enumeration procedures
- Agricultural seasonality in survey timing
- Resource constraints and cost-efficiency measures

Contact Information

Workshop Coordinator: Dr. Harry [Surname]

Technical Support: [sadc.sampling.support@\[domain\]](mailto:sadc.sampling.support@[domain])

Repository Maintenance: github.com/SADC-Stats/sampling-workshop

License

Materials developed under Creative Commons Attribution 4.0 International License (CC BY 4.0) with acknowledgment to Eurostat, World Bank, OECD, and SADC Secretariat.

Acknowledgments

This workshop was developed with technical assistance from:

- European Statistical System (ESS)
 - World Bank Living Standards Measurement Study (LSMS) Team
 - OECD Programme for the International Assessment of Adult Competencies (PIAAC)
 - SADC Statistics Unit
 - National Statistics Offices of all 16 SADC Member States
-

Last Updated: [Current Date]

Version: 1.0.0